



Energy Bilocalization Effect and the Emergence of Molecular Functions in Proteins

Yann Chalopin* and Julien Sparfel

Laboratoire EM2C–CNRS and CentraleSupélec, University of Paris-Saclay, Gif-sur-Yvette, France

Proteins are among the most complex molecular structures, which have evolved to develop broad functions, such as energy conversion and transport, information storage and processing, communication, and regulation of chemical reactions. However, the mechanisms by which these dynamical entities coordinate themselves to perform biological tasks remain hotly debated. Here, a physical theory is presented to explain how functional dynamical behavior possibly emerge in complex/macro molecules, thanks to the effect that we term bilocalization of thermal vibrations. More specifically, our approach allows us to understand how structural irregularities lead to a partitioning of the energy of the vibrations into two distinct sets of molecular domains, corresponding to slow and fast motions. This shape-encoded spectral allocation, associated to the genetic sequence, provides a close access to a wide reservoir of dynamical patterns, and eventually allows the emergence of biological functions by natural selection. To illustrate our approach, the SPIKE protein structure of SARS-COV2 is considered.

OPEN ACCESS

Edited by:

Marco Pasi,
École normale supérieure Paris-
Saclay, France

Reviewed by:

Luisa Di Paola,
Campus Bio-Medico University, Italy
Yves-henri Sanejouand,
Unité de fonctionnalité et Ingénierie
des Protéines (UFIP), France

*Correspondence:

Yann Chalopin
yann.chalopin@cns.fr

Specialty section:

This article was submitted to
Biological Modeling and Simulation,
a section of the journal
Frontiers in Molecular Biosciences

Received: 05 July 2021

Accepted: 20 October 2021

Published: 23 December 2021

Citation:

Chalopin Y and Sparfel J (2021) Energy
Bilocalization Effect and the
Emergence of Molecular Functions
in Proteins.
Front. Mol. Biosci. 8:736376.
doi: 10.3389/fmolb.2021.736376

Keywords: protein, rate promoting vibrations, dynamics, vibrations, localization

INTRODUCTION

In his seminal book “What is Life,” Erwin Schrödinger exposed his physical vision of modern molecular biology by priorly emphasizing the intrinsically aperiodic character of living microscopic systems (notably the nucleic acids). His vision arose from the observation that such biological structures failed the physical theories of the time, which were centered, for the most part, on the principle of the periodicity of crystal structures. Indeed, it was the introduction of the concept of regularity (Gross, 1996) (or symmetries) that supported the development of a deep understanding of the laws of quantum mechanics when they are applied to describe matter, its electronic and thermal properties, and its relationship with radiation. Heat in solids corresponds to crystalline vibrations defining quasi-particles called phonons (it is necessary to consider electrons in metals). Phonon transport (heat flux) corresponds to the formation of wave packets that are subjected to dispersion phenomena and propagate according to a mean-free path controlled by various relaxation processes specific to the crystal structure. Heat capacity, thermal expansion or even thermal conductivity are now quantities predicted with substantial precision (Ashcroft and David Mermin, 1976), thanks to the well known physical concept of first Brillouin zone.

Conversely, the transport of energy and/or information in living systems is dominated by intricate cascades of molecular interactions orchestrated by completely irregular structures called proteins. Proteins are macromolecules derived from the translation of RNA into polymer chains of amino-acids (AA), which fold to form unique and irregular structures (Hay and Scrutton, 2012). To operate, proteins need motions. More specifically, this is achieved by the thermal environment: heat constantly

maintains proteins dynamics through thermal atomic displacement fluctuations from the thermal bath (in most cases, the solvent), at timescales ranging from the fs (e.g., proton transfer) to the ms timescales (e.g., conformational changes) (Damry et al., 2019). This dynamics plays for instance a determining role in enzyme catalysis (Wolf-Watz et al., 2004; Yang and Bahar, 2005; Henzler-Wildman et al., 2007; Boekelheide et al., 2011; Callender and Dyer, 2015; Keul et al., 2018), allostery (del Sol et al., 2007), and molecular recognition. What we want to stress here is that a protein is an object that results from a regular system's distortion towards a disordered/inhomogeneous structure or topology, because we believe that the coexistence between structural order and disorder produces dynamical patterns that have been perceived as functional throughout natural selection (Keul et al., 2018; Borgia et al., 2018). Why could it be so? Despite the demonstrated importance of protein dynamics, the interplay between the genetic sequence and the time sequence of functional motions remains widely undiscovered (Panca et al., 2016). Current efforts in contemporary molecular biology focus particularly on finding ways to describe proteins as structure-driven dynamical entities. While it has long been a question of how the structure dictates the function, the current paradigm should be shifted towards quantitatively describing function more as a result of dynamical properties, which are themselves driven by shapes (Ishima and Torchia, 2000). Unlike solids, we claim here from the basis of mathematical arguments that living organisms ultimately rest on the ability of evolution to shape irregular molecular structures (Keul et al., 2018) to produce functional dynamical patterns. Unfortunately, developments in molecular biology and condensed matter physics have made their way apart. Consequently, few studies have apprehended proteins' dynamics with a starting point the entanglement between disorder (shape) and propagation phenomena sustained by the environment (heat). The transport of non-interacting electrons in disordered structure has been extensively investigated (Filoche and Mayboroda, 2012; Arnold et al., 2016; Lemut et al., 2020) [the main known effect of which is Anderson localization (Anderson, 1958)] while very few works have formalized localization phenomena in phonon systems (Keppens et al., 1998; Seyf et al., 2017; Juntunen et al., 2019; Zhou et al., 2020). In this letter, we are inspired by a powerful mathematical tool (Filoche and Mayboroda, 2012) to describe a theory which offers the possibility to understand the entanglement between microscopic lattice vibrations and propagation phenomena in a complex topology, in order to give meaning to the dynamics of proteins and enzymes. The following study is based on a central assumption, which is that such dynamical systems, potentially functional, arise from sets of motions involving the largest number of different domains, themselves associated to the most distinct characteristic times and amplitudes to be selected (McDonald et al., 2013). This leads us to address proteins' functioning by formulating a question in physical terms: how

to reconstruct the functional dynamics -that is which linear combination of normal modes-of a macro-molecule by associating each characteristic time to a spatial domain, from the knowledge of the 3D structure solely? Finally, once this quantity is mapped on proteins' topology, what is revealed?

MATERIALS AND METHODS

A General Description of the Protein Dynamics

Any periodic sequence composed of invariantly coupled degrees of freedom universally leads to dispersive wave transport. Like any other folded polymer system, a protein can be first approximated as a linear chain of AA bonded between the C_α atoms (backbone) as shown on **Figure 1**. However, these covalent interactions are completed by noncovalent bonds between the side chains: proteins fold, and consequently, each degree of freedom of the BB has cohesive energy that is no longer spatially homogeneous **Figure 1**. Based on these physical considerations, we express the lagrangian of the system as,

$$\mathcal{L} = \sum_{i=1}^N \frac{1}{2} \dot{X}_i^2 - \sum_{1 \leq i < j \leq N} \frac{1}{2} \alpha_{ij} (X_i - X_j)^2 \quad (1)$$

which allows to extract the harmonic equations of motion of such a dynamical system as

$$\omega^2 X_i = \left(\sum_{j=1}^N \alpha_{ij} \right) X_i - \sum_{j=1}^N \alpha_{ij} X_j \quad (2)$$

where ω stands for the harmonic pulsation, X_i is the displacement of atom i from its equilibrium position x_i . The α_{ij} are constants characterizing the strength of the bond between atoms i and j . We see from this eq. that disorder (irregularity) is reflected by the quantities α_{ij} which account for the nature of the interactions (ex. angular or distant dependent) as well as the number of "neighbors" included in the sum. We introduce the matrix (operator):

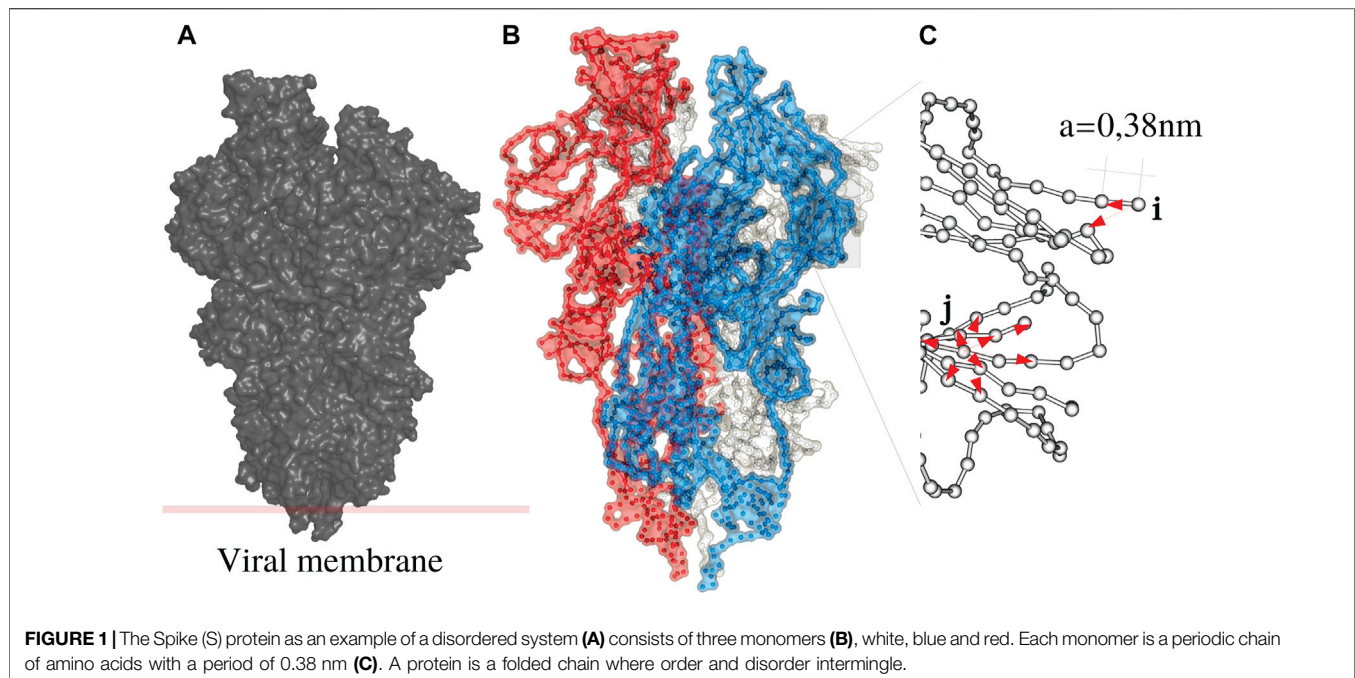
$$-\Delta : (\delta_{ij}) = \begin{cases} \sum_{k=1}^N \alpha_{ik}, & \text{if } i = j. \\ -\alpha_{ij}, & \text{otherwise.} \end{cases} \quad (3)$$

such that the equations of motion are casted into an eigenproblem in the form of a Laplace equation:

$$-\Delta X = \omega^2 X \quad (4)$$

We consider vibrations defined at the zone-edge, by the wavenumber $k = \pi/a$: $\tilde{X}_j = e^{i\pi x_j/a} X_j = (-1)^j X_j$. We introduce a real constant C , taken slightly greater than the greatest eigenvalue ω_{\max}^2 of $-\Delta$ and write

$$(C - \omega^2) \tilde{X}_i = \left(C - \sum_{j=1}^N \alpha_{ij} \right) \tilde{X}_i + \sum_{j=1}^N \alpha_{ij} (-1)^{j-i} \tilde{X}_j \quad (5)$$



with an operator L_h defined by:

$$L_h \tilde{X} = (C - \omega^2) \tilde{X} \quad (6)$$

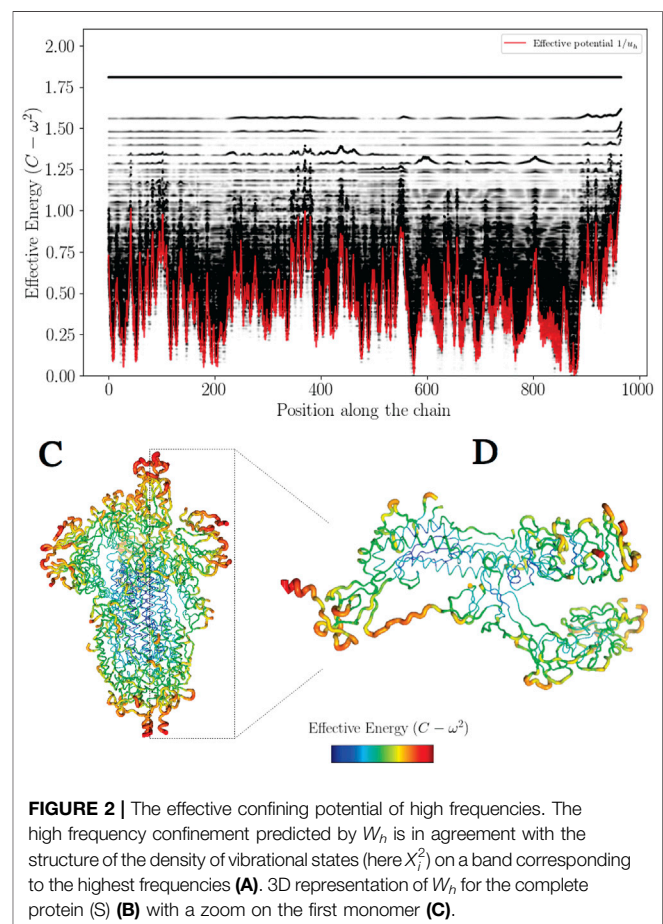
Localized Fast and Slow Motions and Bilocalization Effects

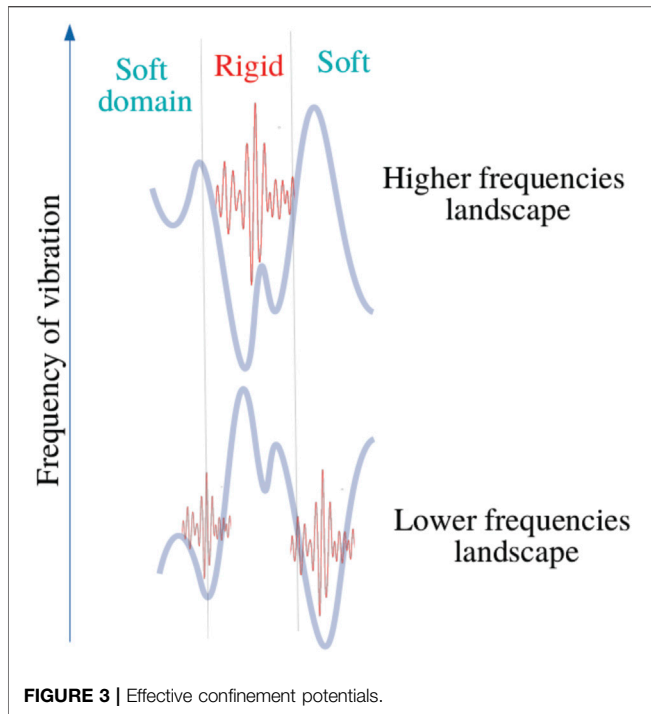
A first frequency localization landscape (LL) u_h is obtained from

$$L_h u_h = (1), \quad (7)$$

where $(1) = (1, \dots, 1)^T$, which is termed high frequency LL for its ability to capture the localization of the highest frequency modes (Chalopin et al., 2019). In the case of a continuous elliptic differential operator H on a domain Ω , it has been shown that the localization landscape u defined as $Hu = 1$ on Ω satisfies $\frac{|\Psi(\mathbf{x})|}{\|\Psi\|_\infty} \leq Eu(\mathbf{x})$ where Ψ is an eigenfunction of H with eigenvalue E . We would then expect in our case to have $\frac{\tilde{X}_i}{\|\tilde{X}\|_\infty} \leq (C - \omega^2)u_i$ for every $1 \leq i \leq N$. Unlike in the continuous case this is not generally true. The maximum principle is indeed not true for the L_h operator. Components of L_h^{-1} are then not strictly positive and the inequality is false in general. However we numerically find that this inequality is fulfilled on most cases. Therefore, high-frequency modes are expected to be strongly localized on the maxima of u_h or equivalently on the minima of $W_h = 1/u_h$.

In a discrete disordered system the dispersion relation is no longer monotonous: wave-vectors at the zone edge are no longer systematically associated with the highest frequencies. For this reason low frequency modes are expected to be localized too. We introduce a complementary localization landscape that captures low frequency localization at the zone edge.





We reconsider the dynamics of the system written as:

$$(\omega^2 + \epsilon)\tilde{X}_i = \left(\epsilon + \sum_{j=1}^N \alpha_{ij} \right) \tilde{X}_i - \sum_{j=1}^N \alpha_{ij} (-1)^{j-i} \tilde{X}_j \quad (8)$$

and introduce the operator L_i :

$$L_i \tilde{X} = (\omega^2 + \epsilon)\tilde{X} \quad (9)$$

which is positive definite with ϵ a small constant. The low frequency localization landscape u_l is straightforwardly obtained from:

$$L_l u_l = (1) \quad (10)$$

As previously the inequality $\frac{\tilde{X}_i}{\|\tilde{X}\|_\infty} \leq (\omega^2 + \epsilon)u_i$ is not generally true but happens to be true on most cases. On **Figure 2** the effective confining potential $W_l = 1/u_l$ (red) of the Spike protein of SARS-Cov2 is displayed. One can clearly see that low frequency modes are localized on the minima of W_l . As it can be seen on **Figure 3**, the 2 localization landscapes happen to be complementary: the minima of the high frequency LL are located on the same sites as the maxima of the low frequency LL and conversely. This can be explained by a simple relation between operators L_h and L_l :

$$L_h = (C + \epsilon)\mathbb{I}_N - L_l \quad (11)$$

where \mathbb{I}_N denotes the $N \times N$ identity matrix. L_l being symmetric and positive definite, there exists an orthogonal matrix O and a diagonal matrix Σ with strictly positive eigenvalues such that:

$$L_l = O\Sigma O^T \quad (12)$$

Note that matrix O can be chosen such that it is the matrix of change of basis from the canonical basis to the basis of envelopes at $k = \pi/a$ (which are eigenvectors of L_l). Localization landscapes are defined as:

$$\begin{cases} L_h u_h = (1) \\ L_l u_l = (1) \end{cases} \quad (13)$$

We now use the diagonalization of L_l in these two expressions:

$$\begin{cases} ((C + \epsilon)\mathbb{I}_N - O\Sigma O^T)u_h = (1) \\ O\Sigma O^T u_l = (1) \end{cases} \quad (14)$$

$$\Rightarrow \begin{cases} (O^T(C + \epsilon)\mathbb{I}_N - \Sigma O^T)u_h = O^T(1) \\ \Sigma O^T u_l = O^T(1) \end{cases} \quad (15)$$

$$\Rightarrow \begin{cases} ((C + \epsilon)\mathbb{I}_N - \Sigma)u_h^* = (1)^* \\ \Sigma u_l^* = (1)^* \end{cases} \quad (16)$$

where we have denoted $v^* = O^T v$ for any column vector v and we have used the fact that O^T and $(C + \epsilon)\mathbb{I}_N$ commute. Thus we have:

$$((C + \epsilon)\mathbb{I}_N - \Sigma)u_h^* = \Sigma u_l^* \quad (17)$$

Matrices Σ and $(C + \epsilon)\mathbb{I}_N - \Sigma$ being diagonals we have for every $1 \leq i \leq N$:

$$(u_h^*)_i = \frac{\Sigma_{ii}}{C + \epsilon - \Sigma_{ii}} (u_l^*)_i$$

where we have denoted $(v)_i$ the i th component of a column vector v . The Σ_{ii} are eigenvalues of matrix L_l so we have $\Sigma_{ii} = \omega_i^2 + \epsilon$ for every $1 \leq i \leq N$. Eventually we obtain a very simple relation between components of the localization landscapes in the basis of envelopes at $k = \pi/a$:

$$(u_h^*)_i = \frac{\omega_i^2 + \epsilon}{C - \omega_i^2} (u_l^*)_i \quad (18)$$

Recall that we choose constant C slightly greater than the greatest eigenvalue ω_{\max}^2 of $-\Delta$. Therefore, for low frequency ω_i **Eq. 18** implies $(u_h^*)_i \ll (u_l^*)_i$. For high frequency ω_i **Eq. 18** implies $(u_h^*)_i \gg (u_l^*)_i$.

How to Look at the Structural Disorder as Quantum-Like Confinement Potentials

Here we describe how the inverse quantities for both landscapes ($1/u$) shall be regarded as effective confining potentials (CP). This important quantity (CP) has been priorly introduced to successfully describe confinement of electronic wavefunctions (Arnold et al., 2016).

We demonstrate that in discrete phonon systems, any localization landscape aforementioned (e.g u_l) also leads to the apparition of an effective potential ($W_l = 1/u_l$) in the equations of motion (**Eq. 25**), which take a form similar to the Schrödinger equation describing a free particle (e.g electron) in an irregular quantum well.

We start with the equation defining the operator L_l :

$$L_l \tilde{X} = (\omega^2 + \epsilon)\tilde{X}, \quad (19)$$

which can be expanded in the continuous limit as

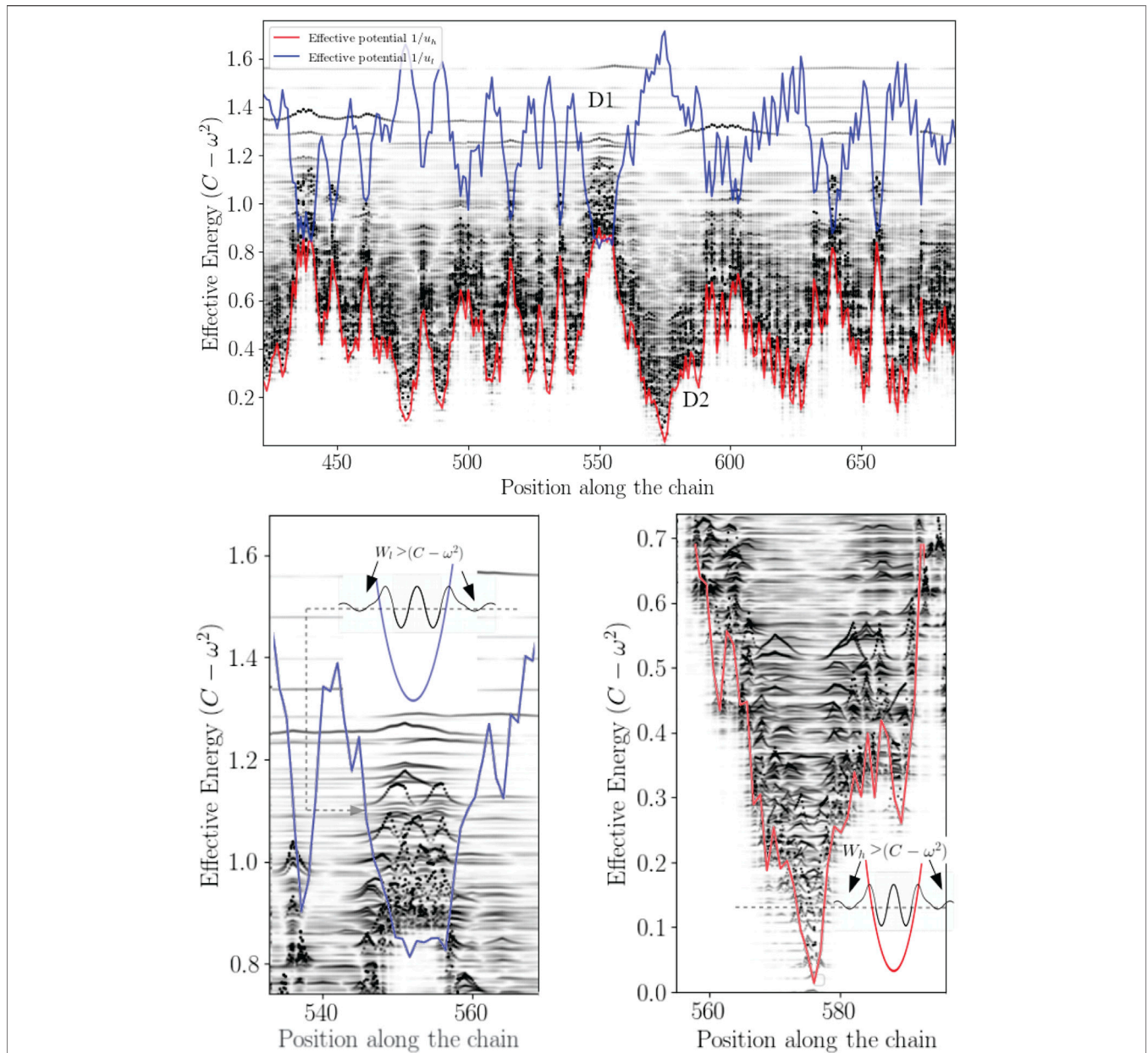


FIGURE 4 | Bilocalization effect The modulus of the normal modes are plotted and shifted according to their eigenvalues by $C - \omega^2$ (dark points). The confinement potentials corresponding to high (red) and low frequency modes (blue) are spatially complementary. The modal amplitude decays exponentially inside the potential.

$$-\text{div}(A\nabla) + V\tilde{X} = (\omega^2 + \epsilon)\tilde{X} \tag{20}$$

We introduce $\tilde{X} = u\phi$ to account for the modulation of the mode with an envelop defined by $u = u_l$. This provides:

$$-\text{div}(Au\nabla\phi) - \text{div}(A\phi\nabla u) + Vu\phi = (\omega^2 + \epsilon)u\phi. \tag{21}$$

Using the Leibniz Formula

$$\text{div}(a\vec{x}) = a\nabla \cdot \vec{x} + \vec{x} \cdot \nabla a \tag{22}$$

we obtain

$$\begin{aligned} & -u\nabla \cdot (A\nabla\phi) - (A\nabla\phi) \cdot \nabla u - \phi\nabla \cdot (A\nabla u) - A\nabla u \cdot \nabla\phi + Vu\phi \\ & = (\omega^2 + \epsilon)u\phi - u\nabla \cdot (A\nabla\phi) - 2A\nabla u \cdot \nabla\phi + \phi(-\nabla \cdot (A\nabla u) \\ & \quad + Vu) \end{aligned}$$

$$= (\omega^2 + \epsilon)u\phi - \nabla \cdot (A\nabla\phi) - 2A \frac{\nabla u}{u} \cdot \nabla\phi - \frac{\phi}{u} = (\omega^2 + \epsilon)\phi \tag{23}$$

where we used the fact that $-\nabla \cdot (A\nabla u) + Vu = 1$. Eventually we get

$$\left[-\frac{\nabla \cdot (u^2 A \nabla)}{u^2} + \frac{1}{u} \right] \phi = (\omega^2 + \epsilon) \phi \quad (24)$$

replacing by the vector X , we obtain

$$-\frac{1}{u_l^2} \operatorname{div}(u_l^2 A \nabla X) + \frac{X}{u_l} = \omega_n^2 X. \quad (25)$$

To illustrate this feature, the high frequency confinement potential has been color-coded in 3D on **Figure 2**.

This analogy leads to introduce the concept of effective barrier through the quantity $W_h - (C - \omega^2)$. Consequently, phonons wavenumbers become purely imaginary where $(C - \omega^2) - W_h < 0$, that is wave amplitudes vanish with exponential decays (**Figure 3** and **Figure 4**). This later aspect is crucial for understanding how a protein operates from the knowledge of its energy partition.

In this work, most the discussion is exemplified on the SPIKE (S) protein of SARS-CoV-2, which is a molecule that plays a key role in the virus replication (Turoňová et al., 2020). Without providing too much details, one can simply recall that it serves as a receptor to bind to the angiotensin 2 converting enzyme of the host receptors, while initiating a binding with the cell membrane before fusing with the latter (S) is formed by three chains (a trimer) of polypeptide polymers (**Figure 1B**). Each chain (monomer) consists of a sequence of amino acids regularly spaced by 0.38 nm (see **Figure 1**) while the folded structure (3D) form a very irregular object. Its dynamical properties (vibrations) play a crucial role in understanding how the virus interacts with its host (Di Paola et al., 2020). The normal modes are calculated from $-\Delta$ considering a parameter free elastic network model (pfENM-GNM) (Yang et al., 2009) as shown in **Figure 3**. The effective high-frequency localization potential obtained from Eq. 7 is also embedded in the figure to highlight confinement effects by the potential: just like Anderson localization, phonon confinement manifests itself by an exponential modal decay upon the effect of an effective barrier. However, W_h does not completely describe the whole spectrum (*i.e.*, low-frequency motions). When the effective energy $C - \omega^2$ rises above the maxima of the potential W_h , novel localizing regions appear where the dynamics were priorly forbidden. Interestingly, each site that is “frozen” within a given frequency band switches to a dynamically active domain, once driven in the complementary frequency band. **Figure 3** and **Figure 4** illustrate this complementary relationship between W_l and W_h as predicted by Eq. 11. These potentials are said reciprocal, and the dynamical result of their interplay is termed bilocalization.

This observation leads to the next question: estimating the density of states (DOS) without solving the eigenvalue problem. Weyl’s law offers an asymptotic estimate of the DOS, and it has been successfully employed in the case of electronic wavefunctions (Arnold et al., 2016) where the number of modes with energy lower than E is estimated as

$$\begin{aligned} N(E) &\approx N_W(E) \\ &= (2\pi)^{-n} \int_{k^2 + W(x) \leq E} 0^n \mathbf{x} 0^n \mathbf{k} \end{aligned} \quad (26)$$

Where W corresponds to the localization landscape for electrons in a random potential and E the effective energy. In our case, since normal modes are localized in space, by Fourier properties their width in k -space is expected to be large. Thus, one may assume that all wave vectors are accessible to a state localized on a scale of few DOF. Instead of summing over the volume of phase space such that $k^2 + W_h(x) \leq C - \omega_0^2$ we sum over all wave vectors of the Brillouin zone and only impose a condition on position: $W_h(x) \leq C - \omega_0^2$. To go from a continuous integral to a discrete sum we use $0k \rightarrow \frac{\pi}{Na}$, $0x \rightarrow a$ and $\int \rightarrow \sum$. The estimation is then:

$$\begin{aligned} N(C - \omega_0^2) &\approx N_{W_h}(C - \omega_0^2) \\ &= \frac{1}{2\pi} \sum_{x_i, k_n} \frac{\pi}{Na} a \\ &\quad W_h(x_i) \leq C - \omega_0^2 \\ &= \frac{1}{2N} \sum_{x_i} \sum_{k_n} 1 \\ &\quad W_h(x_i) \leq C - \omega_0^2 \\ &= \frac{1}{2N} \sum_{x_i} 2N \\ &\quad W_h(x_i) \leq C - \omega_0^2 \\ &= \sum_{x_i} 1 \\ &\quad W_h(x_i) \leq C - \omega_0^2 \end{aligned} \quad (27)$$

$N(x)$ corresponds to the number of modes having a frequency such that $C - \omega^2 \leq x$ and should be termed the eigenvalue counting function (Arnold et al., 2016). **Figure 5** shows comparison of this estimation and the true counting function $N(C - \omega_0^2)$ for (S).

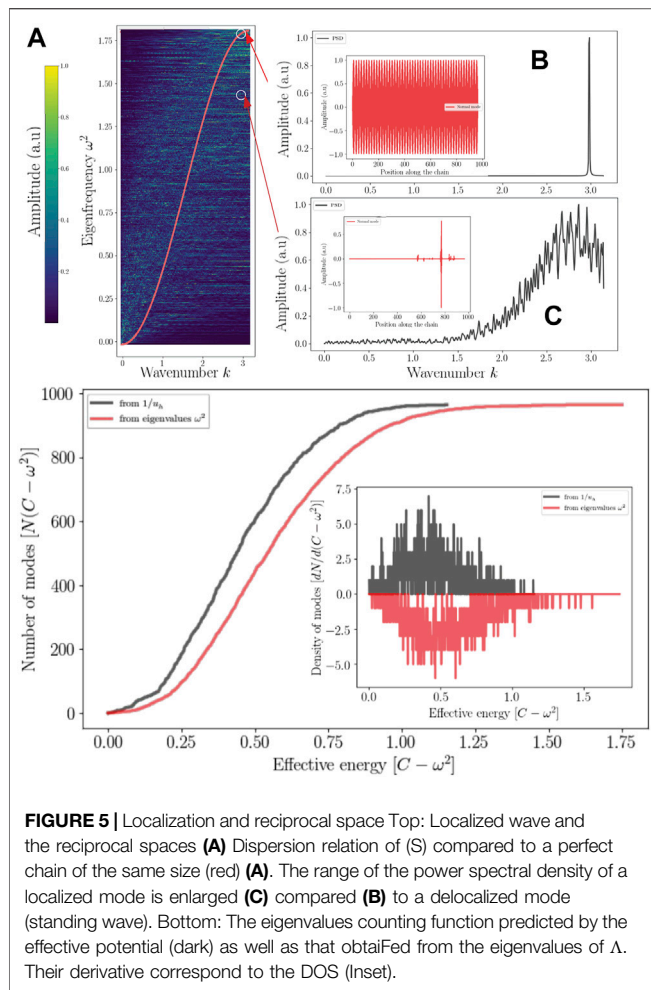
Eq. 27 tends to overestimate the number of high-frequency modes but gives a fair approximation of the density of states. What emerges from these results, which can be generalized for any discrete disordered dynamical system, is that.

1–Disordering a molecular crystal produces two localized energy bands predicted by a double wave confinement potential. These potentials have the particularity of being complementary in the sense that (Eq. 1) temporally, they predict the locations of slow (u_l) and fast (u_h) motions, moreover (2) the domains associated with each temporally distinct landscapes are spatially complementary: their association reproduces the molecular domain as a whole (**Figure 4**).

2–The localization potential of slow motions is a reading of the local rigidity (see Sup. Mat.). Thus, any mode of effective energy higher than the quantity $\sum \alpha_{ij}$ is no longer influenced by W_l . Above a threshold corresponding to the maximum of the W_b , the localization potential W_h takes over.

3–The localized modes are subject to uncertainty $\Delta x \cdot \Delta k \geq 2\pi$, so any strong localization of the energy of a mode implies a broadening of the spectrum associated with the spatial frequencies. Disorder effect on dispersion can be approximated by the modes’ projection at the edge of the BZ. **Figure 5** illustrates this aspect.

4–Complementarity implies that whatever the characteristic time considered, for each DOF, there is always a cutoff frequency below which the dynamics will be inhibited and vice-versa. The



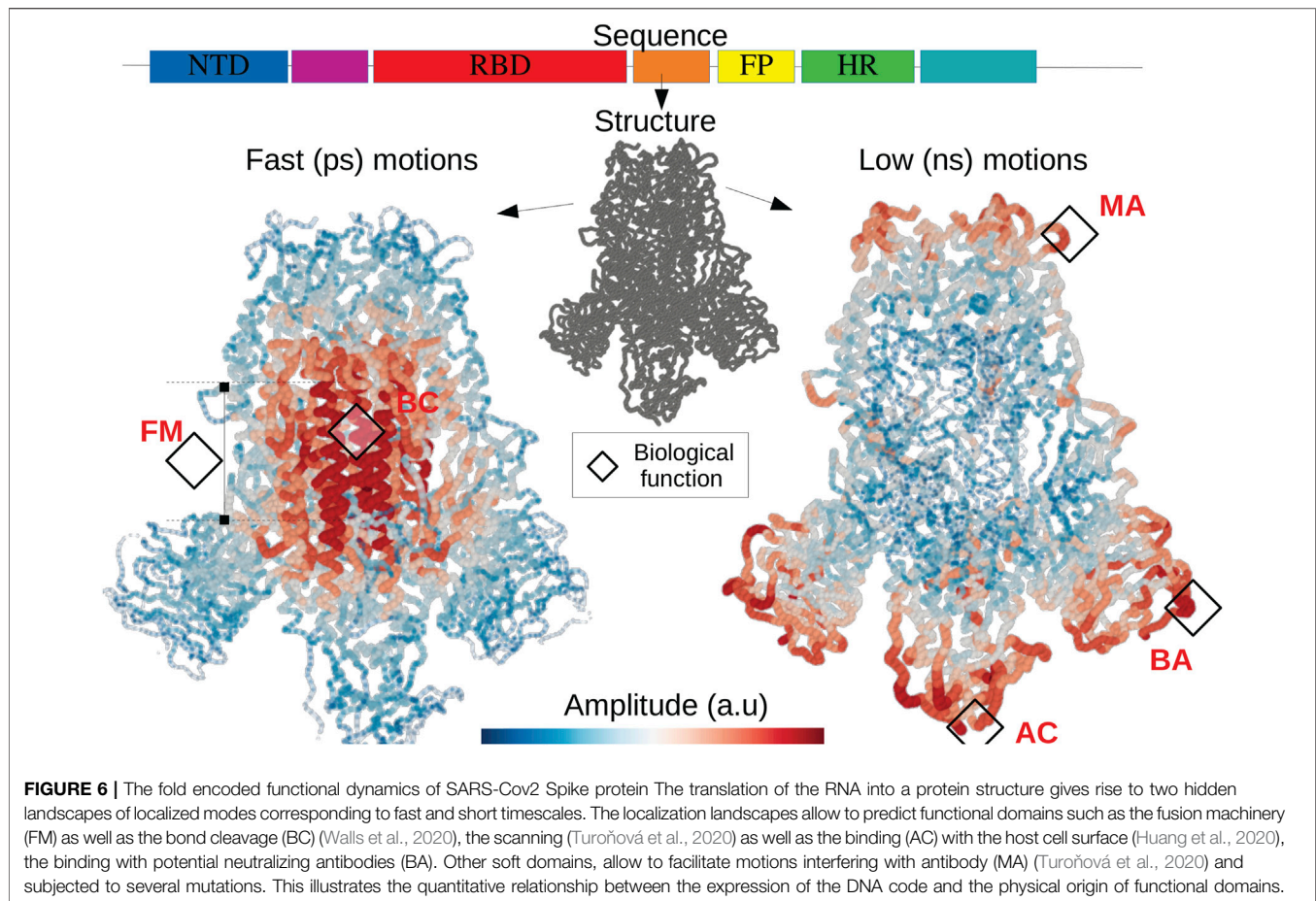
concept of bi-localization implies that there are no frozen domains. Being active or at rest, this only occurs on a limited frequency band. This description is consistent with the equipartition theorem of energy.

DISCUSSION

Whether a many-body disordered system can be thermalized is a question that has remained unanswered for a long time (Wang et al., 2020). We show here that any coherent transport regime, integrated into a discrete disordered structure, produces two distinct broadband localization effects similar to Anderson's localization but contrary to the case of non-interacting electrons (Lyra et al., 2015), strong localization occurs at the zone edge for high and low energies. Such partition of energy spectrum into two spatially complementary subdomains avoids the freezing of a DOF when the system is set in contact with a thermostat. Therefore bilocalization establishes a prerequisite for energy equipartition in phonon systems. In other words, discrete disordered systems at the equipartition produce a rich dynamics, *i.e.*, involving motions in many subdomains and this at very different characteristic times. These two

localization effects imply a broadening of the spatial frequency spectrum resulting in zero group-velocity modes (Arregui et al., 2019), which dramatically reduces the heat flux. Although this subject is out of the scope of the paper, the reduction in thermal conductivity observed in disordered solids (Keppens et al., 1998; Seyf et al., 2017; Juntunen et al., 2019; Zhou et al., 2020) must therefore be understood as the consequence of a projection of the phonon dispersion relation at the zone-edge. Nature, for its part, can take advantage of such dynamical effects associated to wave localization. Proteins appear to be thermally driven nanomachines: slow modes at the *ns* to *ms* timescales (also termed conformational dynamics) correspond to conformational changes of large molecular domains, they are of fundamental importance and therefore extensively studied (Bahar et al., 1999; Henzler-Wildman et al., 2007; Callender and Dyer, 2015; Damry et al., 2019). Their biological functions often lie in the establishment of characteristic times associated with the regulation of chemical reactions such as enzymatic catalysis (Wolf-Watz et al., 2004) as well as allostery (del Sol et al., 2007). Large scale dynamics is well predicted by the landscape W_l . The application of this framework of analysis on (S) allows highlighting the flexibility at the level of contact with the viral membrane, which allows the latter to orient itself effectively to scan the host cell surface as well as to bind ACE2 (Turoňová et al., 2020). The conformational movements called “up” and “down” are also well described (Huang et al., 2020; Walls et al., 2020) by the landscape W_l . The faster functional motions (on the ps time range) are termed rate-promoting vibrations (RPV) (Henzler-Wildman et al., 2007; Hay et al., 2010; Hay and Scrutton, 2012; Schramm and Schwartz, 2018). They play the role of increasing the kinetics of chemical reactions at the level of active sites. Their physical origin has been extensively debated (Schramm and Schwartz, 2018; Antoniou and Schwartz, 2001). The geography of hot-spots hosting fast localized motions is very well reproduced by the W_h potential (Chalopin et al., 2019). For (S), these hot spots reveal cavities corresponding to the fusion machinery. The bond cleavage occurring at the fusion-peptide region (Walls et al., 2020) collocates with a domain of fast compressive vibrations (as seen on Figure 6 in red) to activate the protein for membrane fusion. All localization hot-spots revealed by both landscapes correspond to functional domains considered as potential therapeutic targets (Di Paola et al., 2020; Huang et al., 2020; Walls et al., 2020). These two-folded localized dynamics, evolving at two distinct timescales, are likely functionally relevant, and at least universally present in enzymes and all proteins in general. For a function to emerge, something has to be selected. If all the DOFs of a molecule, such as a non-folded polymer, move coherently in time and space, then there is no machinery to select. It is the association of as many frequencies as possible with different domains that produces the richest dynamics. Bilocalization can therefore be perceived as a way to increase the variability of the distribution of these characteristic times over different domains of the molecule.

This demonstrates that the coexistence of order and disorder tends to produce systems capable of evolving with an enhanced



variability of dynamical patterns, offering to evolution a possibility to achieve molecular functionalities by selecting the 3D topology. Is that all ?

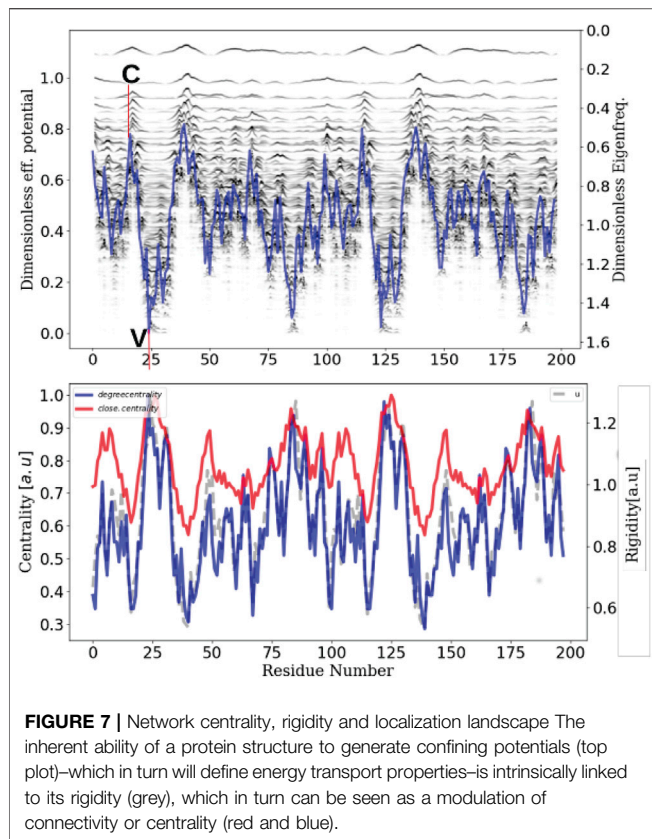
This particular description of the energetic properties exposed here must be further discussed and confronted with previous studies. It is therefore appropriate to present in particular the work of Leitner et al. (2015) with whom we share this idea that the origin of the functional properties of proteins and enzymes is intrinsically associated with the way in which a given structure is able to diffuse heat in the molecular scaffold (Leitner and Yamato, 2020). His approach consists in calculating the local (i.e., residue scale) diffusion on the basis of atomistic simulations such as molecular dynamics. The considerable value that Leitner brings to this problem lies in the ability to relate the transport coefficients (Yu and Leitner, 2005) to the reaction rates by the so-called Master equation (Leitner et al., 2015). We quote here in a non-exhaustive way some works that resonate perfectly with our description. First of all, we fully agree that in these systems, the diffusion of energy does not take place homogeneously, but rather along distinct paths connecting sub-regions of the protein (what we call here the valleys of confinement potential) giving rise to a transport by percolation (Leitner and Yamato, 2020; Leitner, 2008). That is, energy can propagate from residue A to residue B if and only if these residues are part of a connex network defined by their

mechanical coupling (cohesive energy). We believe that hoping transport occurs between different localization valleys through a universal phenomenon known as tunneling effect by evanescent waves. Secondly, we share the conclusion that the transport is anisotropic (Leitner, 2008) since these localization spots are not uniformly distributed in the protein.

Furthermore, Leitner's works also showed that energy percolation between a catalytic residue and other distant residues allows the modulation of chemical bonds through thermal fluctuations (which results in the fluctuations in the length of hydrogen bonds) and that the corresponding network of connected residues correlate well with the mapping of energy flow (Leitner and Yamato, 2020).

We also agree that the role of contacts is preponderant (Leitner et al., 2019; Yamashita et al., 2018). As shown in our model, re-strengthening the cohesive energy locally induces the formation of hot spots (Chalopin, 2020) which can be used as a "hub" (see for instance the literature dealing with graph centrality).

To demonstrate the adequacy between these well established results and our fundamental description of these transport phenomena, we have performed non equilibrium molecular dynamics simulations using the velocity verlet technics coupled to a pFENM hessian (ANM) (Yang et al., 2009). To simplify our discussion as much as possible without diminishing



the general scope of our remarks, we have chosen a much smaller system than SPIKE trimer such as the HIV1-protease enzyme (PDB id:1rx7) and calculate the confinement potential of the dimer. The corresponding plot (top of **Figure 7**) allows us to identify the valleys of the confinement potential in which the energy must remain confined or a contrario the collars which will allow the isolation of the high frequency domains. We have chosen two sites (a collar C) and a valley (V, corresponding to the active sites) and we parametrically excite an AA in each of these domains, at the fastest frequency of the system (here 3 THz).

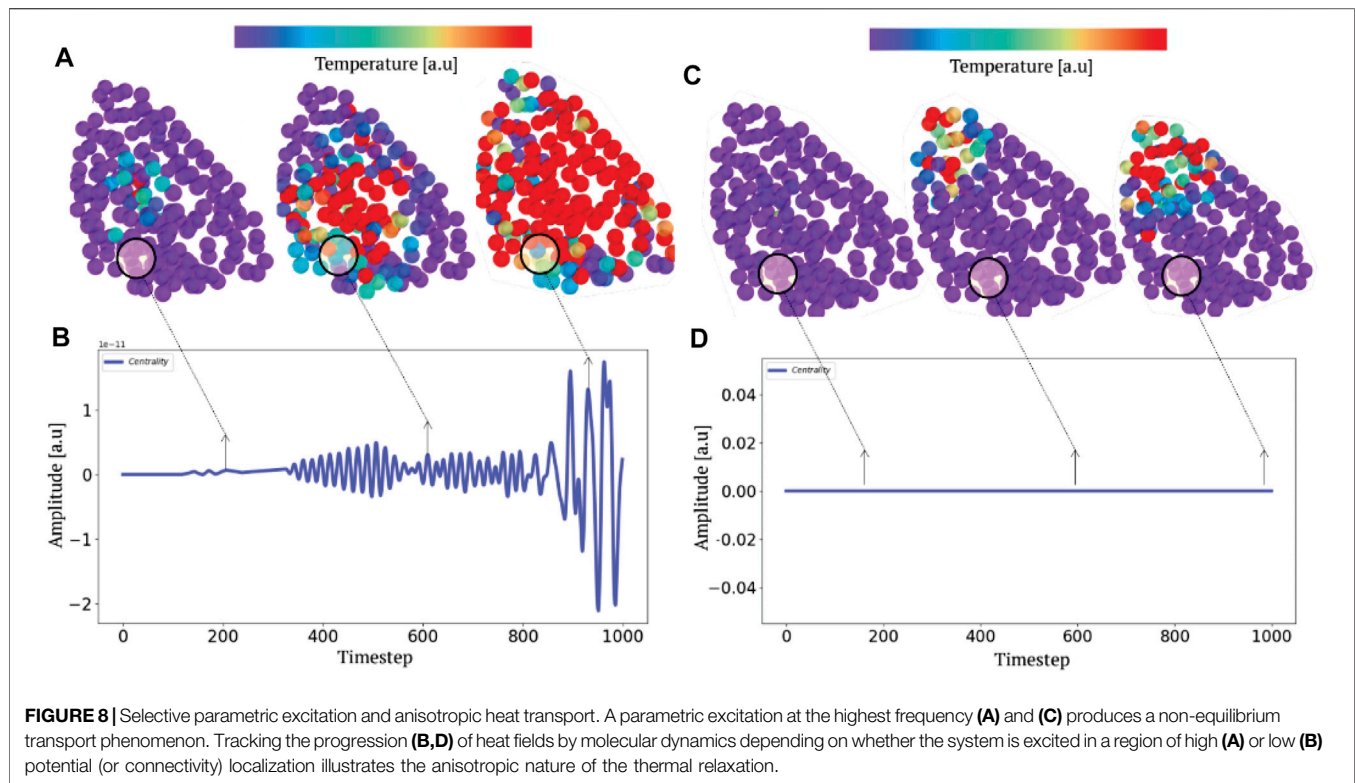
From this selective excitation on the AAs, we follow temporally how the locally injected energy propagates in the scaffold of the protein (**Figure 8**). On the first plot (right), we notice that when we excite the system in a domain that does not correspond to a confinement potential well, the energy does not propagate in the system but remains localized at the level of the excitation (C and D). However, when this excitation is performed at the level of the active sites (left), which correspond to the most confining domains, we notice that the energy propagates in a dramatically more efficient way in the whole system (A and B). This numerical experiment allows to emphasize unequivocally that the transport is indeed anisotropic and that this anisotropy is prescribed by the localization landscape, thus generalizing Leitner's results under a broader physical picture. In order to compare our approach to the state of the art, we calculated the degree of centrality of the system and compared it to the localization

landscape (bottom of **Figure 7**), thus revealing a correlation factor between the two quantities that amounts to 0.9. This last result allows us to claim that rigidity, connectivity or degree of centrality are quantities that relate to the same fundamental physical effect, that is the vibrational confinement potential. Thus, it becomes possible to associate these quantities to strongly anisotropic energy transport properties, and that this is this particular transport mechanism (associated to the percolation through confined subregions) governs the possibility of having locally various functional dynamics.

However, there is an additional concept often recalled in the literature to describe the diffusion of energy which is the mean free path. We believe, on the basis of the physical description presented here, (i.e., that a protein can be seen as a set of potentials wells) that this particular vision of transport is erroneous and here are the reasons for this. The mean free path considers that energy is exchanged through quasiparticles propagating for a certain time or along a certain distance before undergoing a collision inducing a loss of phase of the wave packet and leading in some cases to energy loss or relaxation mechanisms. This view implies that the size of the system is very large compared to the coherence length of the wave packets describing these quasiparticles. We have shown here that the localization is an interferential phenomenon, where the waves propagating in the complex medium will create reflection/transmission phenomena giving rise to the localization spots. Clearly in this type of regime, the coherence length of the waves is of the order of the size of the system otherwise no localization phenomenon could be observed. We therefore believe that the transport in the protein is not a classical scattering phenomenon but rather a hopping transport phenomenon, induced by the set of weak or resonant couplings between localizing subdomains resulting from wave interferences.

Summary of the Main Results

Bilocalization, as it is presented here, is a theory that allows to describe in a condensed and quite robust way the dynamics of proteins, that is the evolution of the positions of the degrees of freedom in space and in time, at the two limits of the slowest and fastest time scales. This tool can be adapted to any cohesion model, atomistic or coarse-grained. If it is well known that a protein is a dynamical object with multiscale behaviors, there are in fact no tools that surpass the traditional normal mode analysis to highlight the domains characterizing the motions at a given frequency/time period. The normal mode approach widely developed in the literature, is neither rigorous nor sufficient to establish a complete picture of proteins and enzymes dynamics. Indeed, computing eigenmodes consists in accessing to the vector space that allows to decompose each motion with a well defined orthonormal basis. In other words, the motion of each atom/AA is a linear combination of the eigenmodes. If it is possible to access to this basis by solving an eigenvalue problem, we learn nothing about the combination of these modes (i.e., we do not know the coefficients that define the linear combination, nor the phase shifts between these modes), thus the information we have access to is partial, and one cannot rigorously predict an observed motion on the



sole knowledge of the eigenvectors. In addition, if we want to study large systems like the SPIKE protein or RNA-polymerase (roughly 5000 AA), it is almost unmanageable to manipulate 15 0,000 eigenvalues with eigenvectors distributed over 50,000 amino acids. In most works, the analysis is often restricted to the few lowest energy mode, which form a too small fraction of the full eigenmodes spectrum. By solving the linear system $LU = 1$, with the right localization operator L , *i.e.*, by extracting a single quantity, associated to each degree of freedom, one gets a more accurate information roughly 100 times faster to calculate because solving a linear system is much easier in terms of algorithmic complexity than calculating the full spectrum of an operator.

The analysis proposed here is not only interesting to predict rigorously the predominant motions observed in these systems, but it also allows to understand the way energy flows in the latter. That is, which domains are energetically coupled and which are isolated (Chalopin et al., 2019).

In addition, bilocalization also offers a deep description of the underlying physical mechanisms that allow to universalize the formation of a partitioning/confinement of vibrations in proteins' domains. To this end, we have established an analogy with quantum mechanics by showing that the structural disorder of the protein, which translates into an inhomogeneity of its rigidity, produces a confinement potential for the phonons (vibration modes) of the system just as a crystal produces a potential barrier for an electron (Filоче et al.). We will see later on, how this physical description allows to unify the set of well known methods such as the analysis of proteins by contact network, the long

range transmission mechanisms induced by rigidity, which are in fact metrics that have some efficiency, but do not share any common physical picture of the exchange processes involved.

CONCLUSION

We have developed a theory to link the topological structure of a molecule (genetically encoded) and its function by studying the (coherent) motions imposed by its complex form. We have shown that structural disorder produces *de facto* a partitioning of its slow and fast motions into several molecular subdomains. Two complementary effective confinement potentials predict these features. At the energy equipartition, each molecular domain is identified by a localized motion of characteristic frequency. This one to one correspondence between topological irregularities and dynamical patterns allows the exploration of large dynamical variability. Regarding the dynamics of a homogeneous collection of harmonic oscillators, disorder/randomness thus appears as a prerequisite to produce a function, that is to say, an efficient temporal and spatial coordination of atomic motions, suited to accomplish a function in a particular environment. This quantitative analysis of this structure-function dynamical interplay forms the basis of a new paradigm for understanding how nature accomplishes molecular design, that is to say, from which generator of properties, evolution manages to select biological functions from an easily modifiable reservoir of dynamical patterns. These localizing properties can be observed experimentally by looking at the rigidity (or its

temperature factor) or by measuring the thermal transport which takes place in these systems and which must appear as highly anisotropic, due to the percolation phenomena produced by the effective confinement potential produced by the disorder.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

REFERENCES

- Anderson, P. W. (1958). Absence of Diffusion in Certain Random Lattices. *Phys. Rev.* 109, 1492–1505. doi:10.1103/physrev.109.1492
- Antoniou, D., and Schwartz, S. D. (2001). Internal Enzyme Motions as a Source of Catalytic Activity: Rate-Promoting Vibrations and Hydrogen Tunneling. *J. Phys. Chem. B* 105, 5553–5558. doi:10.1021/jp004547b
- Arnold, D. N., David, G., Jerison, D., Mayboroda, S., and Filoche, M. (2016). Effective Confining Potential of Quantum States in Disordered media. *Phys. Rev. Lett.* 116, 056602. doi:10.1103/PhysRevLett.116.056602
- Arregui, G., Lanzillotti-Kimura, N. D., Sotomayor-Torres, C. M., and García, P. D. (2019). Anderson Photon-Phonon Colocalization in Certain Random Superlattices. *Phys. Rev. Lett.* 122, 043903. doi:10.1103/PhysRevLett.122.043903
- Ashcroft, Neil. W. (1976). in *Solid State Physics/Neil W. Ashcroft*. Editor N. David Mermin (Philadelphia (Pa.): Saunders College).
- Bahar, I., Erman, B., Jernigan, R. L., Atilgan, A. R., and Covell, D. G. (1999). Collective Motions in Hiv-1 Reverse Transcriptase: Examination of Flexibility and Enzyme Function. *J. Mol. Biol.* 285, 1023–1037. doi:10.1006/jmbi.1998.2371
- Boekelheide, N., Salomón-Ferrer, R., and Miller, T. F. (2011). Dynamics and Dissipation in Enzyme Catalysis. *Proc. Natl. Acad. Sci.* 108, 16159–16163. doi:10.1073/pnas.1106397108
- Borgia, A., Borgia, M. B., Bugge, K., Kissling, V. M., Heidarsson, P. O., Fernandes, C. B., et al. (2018). Extreme Disorder in an Ultrahigh-Affinity Protein Complex. *Nature* 555, 61–66. doi:10.1038/nature25762
- Callender, R., and Dyer, R. B. (2015). The Dynamical Nature of Enzymatic Catalysis. *Acc. Chem. Res.* 48, 407–413. doi:10.1021/ar5002928
- Chalopin, Y., Piazza, F., Mayboroda, S., Weisbuch, C., and Filoche, M. (2019). Universality of Fold-Encoded Localized Vibrations in Enzymes. *Sci. Rep.* 9, 12835. doi:10.1038/s41598-019-48905-8
- Chalopin, Y. (2020). The Physical Origin of Rate Promoting Vibrations in Enzymes Revealed by Structural Rigidity. *Sci. Rep.* 10, 17465. doi:10.1038/s41598-020-74439-5
- Damry, A. M., Mayer, M. M., Broom, A., Goto, N. K., and Chica, R. A. (2019). Origin of Conformational Dynamics in a Globular Protein. *Commun. Biol.* 2, 433. doi:10.1038/s42003-019-0681-2
- del Sol, A., Araújo-Bravo, M. J., Amorós, D., and Nussinov, R. (2007). Modular Architecture of Protein Structures and Allosteric Communications: Potential Implications for Signaling Proteins and Regulatory Linkages. *Genome Biol.* 8, R92. doi:10.1186/gb-2007-8-5-r92
- Di Paola, L., Hadi-Alijanvand, H., Song, X., Hu, G., and Giuliani, A. (2020). The Discovery of a Putative Allosteric Site in the Sars-Cov-2 Spike Protein Using an Integrated Structural/dynamic Approach. *J. Proteome Res., ACS J. Proteome Res.* 19, 0c00273. doi:10.1021/acs.jproteome.0c00273
- Filoche, M., and Mayboroda, S. (2012). Universal Mechanism for anderson and Weak Localization. *Proc. Natl. Acad. Sci.* 109, 14761–14766. doi:10.1073/pnas.1120432109
- Gross, D. J. (1996). The Role of Symmetry in Fundamental Physics. *Proc. Natl. Acad. Sci.* 93, 14256–14259. doi:10.1073/pnas.93.25.14256
- Hay, S., Johannissen, L. O., Sutcliffe, M. J., and Scrutton, N. S. (2010). Barrier Compression and its Contribution to Both Classical and Quantum Mechanical Aspects of Enzyme Catalysis. *Biophysical J.* 98, 121–128. doi:10.1016/j.bpj.2009.09.045
- Hay, S., and Scrutton, N. S. (2012). Good Vibrations in Enzyme-Catalysed Reactions. *Nat. Chem* 4, 161–168. doi:10.1038/nchem.1223
- Henzler-Wildman, K. A., Thai, V., Lei, M., Ott, M., Wolf-Watz, M., Fenn, T., et al. (2007). Intrinsic Motions along an Enzymatic Reaction Trajectory. *Nature* 450, 838–844. doi:10.1038/nature06410
- Huang, Y., Yang, C., Xu, X.-f., Xu, W., and Liu, S.-w. (2020). Structural and Functional Properties of Sars-Cov-2 Spike Protein: Potential Antivirus Drug Development for Covid-19. *Acta Pharmacol. Sin* 41, 1141–1149. doi:10.1038/s41401-020-0485-4
- Ishima, R., and Torchia, D. A. (2000). Protein Dynamics from Nmr. *Nat. Struct. Biol.* 7, 740–743. doi:10.1038/78963
- Juntunen, T., Vänskä, O., and Tittonen, I. (2019). Anderson Localization Quenches thermal Transport in Aperiodic Superlattices. *Phys. Rev. Lett.* 122, 105901. doi:10.1103/PhysRevLett.122.105901
- Keppens, V., Mandrus, D., Sales, B. C., Chakoumakos, B. C., Dai, P., Coldea, R., et al. (1998). Localized Vibrational Modes in Metallic Solids. *Nature* 395, 876–878. doi:10.1038/27625
- Keul, N. D., Oruganty, K., Schaper Bergman, E. T., Beattie, N. R., McDonald, W. E., Kadirvelraj, R., et al. (2018). The Entropic Force Generated by Intrinsically Disordered Segments Tunes Protein Function. *Nature* 563, 584–588. doi:10.1038/s41586-018-0699-5
- Leitner, D. M., Buchenberg, S., Brettel, P., and Stock, G. (2015). Vibrational Energy Flow in the Villin Headpiece Subdomain: Master Equation Simulations. *J. Chem. Phys.* 142, 075101. doi:10.1063/1.4907881
- Leitner, D. M. (2008). Energy Flow in Proteins. *Annu. Rev. Phys. Chem.* 59, 233–259. doi:10.1146/annurev.physchem.59.032607.093606
- Leitner, D. M., Pandey, H. D., and Reid, K. M. (2019). Energy Transport across Interfaces in Biomolecular Systems. *J. Phys. Chem. B* 123, 9507–9524. doi:10.1021/acs.jpcc.9b07086
- Leitner, D. M., and Yamato, T. (2020). Recent Developments in the Computational Study of Protein Structural and Vibrational Energy Dynamics. *Biophys. Rev.* 12, 317–322. doi:10.1007/s12551-020-00661-0
- Lemut, G., Pacholski, M. J., Ovdar, O., Grabsch, A., Tworzydło, J., and Beenakker, C. W. J. (2020). Localization Landscape for Dirac Fermions. *Phys. Rev. B* 101, 081405. doi:10.1103/physrevb.101.081405
- Lyra, M. L., Mayboroda, S., and Filoche, M. (2015). Dual Landscapes in anderson Localization on Discrete Lattices. *Epl* 109, 47001. doi:10.1209/0295-5075/109/47001
- McDonald, L. R., Whitley, M. J., Boyer, J. A., and Lee, A. L. (2013). Colocalization of Fast and Slow Timescale Dynamics in the Allosteric Signaling Protein CheY. *J. Mol. Biol.* 425, 2372–2381. doi:10.1016/j.jmb.2013.04.029
- Panca, R., Raimondi, D., Cilia, E., and Vranken, W. F. (2016). Early Folding Events, Local Interactions, and Conservation of Protein Backbone Rigidity. *Biophysical J.* 110, 572–583. doi:10.1016/j.bpj.2015.12.028
- Schramm, V. L., and Schwartz, S. D. (2018). Promoting Vibrations and the Function of Enzymes. Emerging Theoretical and Experimental Convergence. *Biochemistry* 57, 3299–3308. doi:10.1021/acs.biochem.8b00201
- Seyf, H. R., Yates, L., Bougher, T. L., Graham, S., Cola, B. A., Detchprohm, T., et al. (2017). Rethinking Phonons: The Issue of Disorder. *Npj Comput. Mater.* 3, 49. doi:10.1038/s41524-017-0052-9

AUTHOR CONTRIBUTIONS

YC has conceived the idea and performed the numerical calculations. YC and JS have developed the derivations, analyzed the data, co-wrote the manuscript and finally agree to its content.

ACKNOWLEDGMENTS

The author (YC) would like to thank M. Filoche for useful discussions at the early stage of this work.

- Turoňová, B., Sikora, M., Schürmann, C., Hagen, W. J. H., Welsch, S., Blanc, F. E. C., et al. (2020). *In Situ* Structural Analysis of Sars-Cov-2 Spike Reveals Flexibility Mediated by Three Hinges. *Science*, eabd5223.
- Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020). Structure, Function, and Antigenicity of the Sars-Cov-2 Spike Glycoprotein. *Cell* 181, 281e6–292. doi:10.1016/j.cell.2020.02.058
- Wang, Z., Fu, W., Zhang, Y., and Zhao, H. (2020). Wave-turbulence Origin of the Instability of Anderson Localization against many-body Interactions. *Phys. Rev. Lett.* 124, 186401. doi:10.1103/PhysRevLett.124.186401
- Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E. Z., and Kern, D. (2004). Linkage between Dynamics and Catalysis in a Thermophilic-Mesophilic Enzyme Pair. *Nat. Struct. Mol. Biol.* 11, 945–949. doi:10.1038/nsmb821
- Yamashita, S., Mizuno, M., Tran, D. P., Dokainish, H., Kitao, A., and Mizutani, Y. (2018). Vibrational Energy Transfer from Heme through Atomic Contacts in Proteins. *J. Phys. Chem. B* 122, 5877–5884. doi:10.1021/acs.jpcc.8b03518
- Yang, L.-W., and Bahar, I. (2005). Coupling between Catalytic Site and Collective Dynamics: A Requirement for Mechanochemical Activity of Enzymes. *Structure* 13, 893–904. doi:10.1016/j.str.2005.03.015
- Yang, L., Song, G., and Jernigan, R. L. (2009). Protein Elastic Network Models and the Ranges of Cooperativity. *Proc. Natl. Acad. Sci.* 106, 12347–12352. doi:10.1073/pnas.0902159106
- Yu, X., and Leitner, D. M. (2005). Heat Flow in Proteins: Computation of thermal Transport Coefficients. *J. Chem. Phys.* 122, 054902. doi:10.1063/1.1830431
- Zhou, W. X., Cheng, Y., Chen, K. Q., Xie, G., Wang, T., and Zhang, G. (2020). Thermal Conductivity of Amorphous Materials. *Adv. Funct. Mater.* 30, 1903829. doi:10.1002/adfm.201903829

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Chalopin and Sparfel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.